BSTA 620: Probability Lecture Notes

Di Shu, PhD

Department of Biostatistics, Epidemiology and Informatics
University of Pennsylvania
and
Center for Pediatric Clinical Effectiveness
Children's Hospital of Philadelphia
Di.Shu@pennmedicine.upenn.edu

8th December 2021

Recap by case study

	drug A	drug B	
D+	a	b	m_1
D-	c	d	m_2
	n_1	n_2	N

- Let $p_1 = a/n_1$ be the estimator for $\pi_1 = P(D + |drug = A)$, and $p_2 = b/n_2$ be the estimator for $\pi_2 = P(D + |drug = B)$
- Odds ratio $OR = \dfrac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$, which can be estimated by

$$\widehat{OR} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

Recap by case study

- Why this estimator makes sense
- 95% confidence interval for $\widehat{OR}$
- Use and misuse of OR
- OR vs. risk ratio (RR)
- Another look at the rare disease condition by examining correlation

Review

Definition 1.1.1: The set, S, of all possible outcomes of a particular experiment is called the **sample space** for the experiment

Definition 1.1.2: An **event** is any collection of possible outcomes of an experiment, that is, any subset of S (including S itself)

Review

Definition 1.2.4: Given a sample space S and an associated sigma algebra B, a **probability function** is a function P with domain B that satisfies

- $P(A) \geq 0$ for all $A \quad B$
- $P(S) = 1$
- If $A_1, A_2, \ldots \quad B$ are pairwise disjoint, then $P(\ {}_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ (countable additivity)

- Items 1-3 are called **Kolmogorov axioms** of probability

The calculus of probabilities (consequences of 1.2.4)

Theorem 1.2.8: IF P is a probability function and A is any set in B, then

- $P(\ ) = 0$ where   is the empty set
- $P(A) \leq 1$
- $P(A^c) = 1 - P(A)$

Review

Theorem 1.2.9: If P is a probability function and A and B are any sets in **B**, then

- $P(B \cap A^c) = P(B) - P(A \cap B)$
- $P(A \quad B) = P(A) + P(B) - P(A \cap B)$
- If $A \quad B$, then $P(A) \leq P(B)$

Review

$$P\left(\bigcup_{i=1}^{K} A_i\right) \leq \sum_{i=1}^{K} P(A_i)$$

- Can see this from Theorem 1.2.9 which implies

$$P(A \cup B) \leq P(A) + P(B)$$

- Bonferroni's inequality tells us that if we make the probability of a type I error on any given comparison α/K, then the FWE will be

$$P\left(\bigcup_{i=1}^{K} \{\text{type I error on test } i\}\right) \leq \sum_{i=1}^{K} \alpha/K = \alpha$$

Review

Number of possible arrangements of size r from n objects

	Without replacement	With replacement
Ordered	$\frac{n!}{(n-r)!}$	n^r
Unordered	$\binom{n}{r}$	$\binom{n+r-1}{r}$

Review

Definition 1.3.2: If A and B are events in S, and $P(B) > 0$, then the **conditional probability of** A **given** B, written as $P(A|B)$, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Review

Theorem 1.3.5 (Bayes's Rule): Let $A_1, A_2, \ldots$ be a partition of the sample space, and let B be any set. Then for each $i = 1, 2, \ldots,$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}$$

Review

Definition 1.3.7: Two events A and B are **statistically independent** if

$$P(A \cap B) = P(A)P(B)$$

Definition 1.3.12: A collection of events $A_1, \ldots, A_n$ are **mutually independent** if for any subcollection $A_{i_1}, \ldots, A_{i_k}$, we have

$$P(\cap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} P(A_{i_j})$$

Review

Theorem 1.3.9: If A and B are independent events, then the following pairs are also independent:

- A and B^c
- A^c and B
- A^c and B^c

- Related If two random variables are independent then functions of those random variables are independent

Review

Definition 1.4.1: A **random variable** is a function from a sample space S into the real numbers

- This is a simplified definition

Review

Definition 1.5.1: the **cumulative distribution function or** *cdf* of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P_X(X \leq x) \quad \text{for all } x$$

Review

Theorem 1.5.10: The following two statements are equivalent:

- The random variables X and Y are identically distributed
- $F_X(x) = F_Y(x)$ for every x

Review

Definition 1.6.1: The **probability mass function (pmf)** of a discrete random variable X is given by

$$f_X(x) = P(X = x) \text{ for all } x$$

Definition 1.6.3: The **probability density function (pdf)** $f_X(x)$ of a continuous random variable X is the non-negative function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \text{ for all } x$$

Review

Theorem 1.6.5: A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if

- $f_X(x) \geq 0$ for all x
- $\sum_x f_X(x) = 1$ (pmf) or $\int_{-\infty}^{\infty} f_X(x)dx = 1$ (pdf)

Review

- Common discrete distribution
  - Bernoulli
  - Binomial
  - Poisson (and Poisson Process)
  - Discrete uniform
  - Geometric
  - Hypergeometric
  - Negative binomial
  - Multinomial

Review

- Common continuous distribution
  - Normal
  - Cauchy
  - Uniform
  - Exponential
  - Gamma
  - Weibull
  - Beta
  - log normal
  - Double exponential
  - Chi-squared, t, F
- And their relations

Review

Theorem 2.1.5: Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Let X and Y be defined as in Theorem 2.1.3. Suppose that $f_X(x)$ is continuous on X and that $g^{-1}(y)$ has a continuous derivative on Y. Then the pdf of Y is given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right| & y \quad Y \\ 0 & otherwise \end{cases}$$

Review

Theorem 2.1.8: Let X have pdf $f_X(x)$ and let $Y = g(X)$, and define X as above. Suppose there exists a partition $A_0, A_1, \ldots, A_k$, of X such that $P(X \quad A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_1(x), \ldots, g_k(x)$, defined on $A_1, \ldots, A_k$, respectively, satisfying

- $g(x) = g_i(x)$ for $x \quad A_i$
- $g_i(x)$ is monotone on A_i
- The set Y $= \{y : y = g_i(x)$ for some $x \quad A_i\}$ is the same for each $i = 1, \ldots, k$, and
- $g_i^{-1}(y)$ has a continuous derivative on Y, for each $i = 1, 2, \ldots, k$. Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| & y \quad \mathsf{Y} \\ 0 & otherwise \end{cases}$$

Review

Theorem 2.1.10 (Probability integral transformation): Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then $Y \quad uniform(0, 1)$, that is, $P(Y \leq y) = y$, $0 < y < 1$

Review

Definition 2.2.1: The **expected value or mean** of a random variable $g(X)$, denoted $\mathrm{E}\{g(X)\}$, is

$$\mathrm{E}\{g(X)\} = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) = \sum_{x \in \mathcal{X}} g(x) P(X = x) & \text{if } X \text{ is discrete} \end{cases}$$

provided that the integral or sum exists. If $\mathrm{E}|g(X)| = \infty$, we say that $\mathrm{E}\{g(X)\}$ does not exist

Review

- Consider $g(X) = X$:

Definition: The **expected value or mean** of a random variable X, denoted $\mathrm{E}(X)$, is

$$\mathrm{E}(X) = \begin{cases} \int_{-\infty}^{\infty} x f_X(x) dx & \text{if } X \text{ is continuous} \\ \sum_{x \in X} x f_X(x) = \sum_{x \in X} x P(X = x) & \text{if } X \text{ is discrete} \end{cases}$$

provided that the integral or sum exists. If $\mathrm{E}|X| = \infty$, we say that $\mathrm{E}(X)$ does not exist

Review

Definition 2.3.1: For each integer n, the nth **moment** of X (or $F_X(x)$), μ_n', is

$$\mu_n' = \mathrm{E}(X^n)$$

The nth **central moment** of X, μ_n, is

$$\mu_n = \mathrm{E}\{(X - \mu)^n\}$$

where $\mu = \mu_1' = \mathrm{E}(X)$

Review

Definition 2.3.2: The **variance** of a random variable X is its second central moment, $\text{Var}(X) = \text{E}[\{X - E(X)\}^2]$. The positive square root of $\text{Var}(X)$ is the **standard deviation** of X

- $\text{Var}(X) = \text{E}(X^2) - \{\text{E}(X)\}^2$

Review

Theorem 2.2.5: Let X be a random variable and let a, b and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

- $\text{E}(ag_1(X) + bg_2(X) + c) = a\text{E}\{g_1(X)\} + b\text{E}\{g_2(X)\} + c$
- If $g_1(x) \geq 0$ for all x, then $\text{E}\{g_1(X)\} \geq 0$
- If $g_1(x) \geq g_2(x)$ for all x, then $\text{E}\{g_1(X)\} \geq \text{E}\{g_2(X)\}$
- If $a \leq g_1(x) \leq b$ for all x, then $a \leq \text{E}\{g_1(X)\} \leq b$

Theorem 2.3.4: If X is a random variable with finite variance, then for any constants a and b,

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

Review

Definition 2.3.6: Let X be a random variable with cdf F_X. The **moment generating function (mgf)** of X (or of F_X), denoted by $M_X(t)$, is

$$M_X(t) = \mathsf{E}(e^{tX})$$

provided that the expectation exists for t in some neighborhood of 0 (i.e. $h > 0$ such that, t $(-h, h)$, $\mathsf{E}(e^{tX})$ exists). If the expectation does not exist in a neighborhood of 0, we say that the mgf does not exist

Review

Theorem 2.3.7: If X has mgf $M_X(t)$, then

$$E(X^n) = M_X^{(n)}(0)$$

where we define

$$M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$$

That is, the nth moment is equal to the nth derivative of $M_X(t)$ evaluated at $t = 0$

Review

Theorem 2.3.11: Let $F_X(x)$ and $F_Y(y)$ be two cdfs all of whose moments exist

- If X and Y have bounded support, then $F_X(u) = F_Y(u)$ for all u if and only if $\mathrm{E}(X^r) = \mathrm{E}(Y^r)$ for all integers $r = 0, 1, 2, \ldots$
- If the mgfs exist and $M_X(t) = M_Y(t)$ for all t in some neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u

Review

Theorem 2.3.12 (Convergence of mgfs): Suppose $\{X_i, i = 1, 2, \ldots\}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that

$$\lim_{i \to \infty} M_{X_i}(t) = M_X(t)$$

for all t in a neighborhood of 0, and $M_X(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all x where $F_X(x)$ is continuous, we have

$$\lim_{i \to \infty} F_{X_i}(x) = F_X(x)$$

- That is, convergence of mgfs to an mgf (for $|t| < h$) implies convergence of the cdfs

Review

Theorem 2.3.15: For any constants a and b, the mgf of the random variable $aX + b$ is given by

$$M_{aX+b}(t) = e^{bt} M_X(at)$$

Review

- **Exponential families** A family of pdfs or pmfs is called an exponential family if it can be expressed as

$$f(x;\theta) = h(x)c(\theta)\exp\left(\sum_{i=1}^{k} w_i(\theta)t_i(x)\right)$$

where $h(x) \geq 0$ and $t_1(x), \ldots, t_k(x)$ are real-valued functions of the observation x (they cannot depend on θ), and $c(\theta) \geq 0$ and $w_1(\theta), \ldots, w_k(\theta)$ are real-valued functions of the possibly vector-valued parameter θ (they cannot depend on x)

Review

Definition 3.5.5: Let $f(x)$ be any pdf. Then for any μ $(-\infty, \infty)$ and any $\sigma > 0$, the family of pdfs $(1/\sigma)f((x-\mu)/\sigma)$, indexed by the parameter (μ, σ), is called the **location-scale family with standard pdf** $f(x)$; μ is called the **location parameter** and σ is called the **scale parameter**

Review

Definition 4.1.1: An n**-dimensional random vector** is a function from a sample space S into R^n, n-dimensional Euclidean space

Review

Definition 4.1.3: Let (X, Y) be a discrete bivariate random vector. Then function $f(x, y)$ from R^2 into R defined by $f(x, y) = P(X = x, Y = y)$ is called the **joint probability mass function or joint pmf** of (X, Y)

- To be clearer, the notation $f_{X,Y}(x, y)$ will be used
- Use the joint pmf to calculate probability of any event:

$$P\{(X, Y) \quad A\} = \sum_{(x,y) \quad A} f(x, y)$$

Review

Theorem 4.1.6: Let (X, Y) be a **discrete bivariate random vector with joint pmf** $f_{X,Y}(x, y)$. Then the **marginal pmfs** of X and Y, $f_X(x)$ and $f_Y(y)$, respectively, are given by

$$f_X(x) = \sum_{y \quad R} f_{X,Y}(x, y)$$

and

$$f_Y(y) = \sum_{x \quad R} f_{X,Y}(x, y)$$

Review

Definition 4.1.10: A function $f(x, y)$ from R^2 into R is called a **joint probability density function or joint pdf** of the continuous bivariate random vector (X, Y) if, for every $A \quad R^2$

$$P((X, Y) \quad A) = \int_A \int f(x, y) dx dy$$

- Expectation of a function of a bivariate continuous random variable

$$\mathrm{E}\{g(X, Y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$$

Review

- The **marginal probability density functions** of X and Y are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy, \quad -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx, \quad -\infty < y < \infty$$

- Any function $f(x, y) \geq 0$, (x, y) R^2 with

$$1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy$$

is the joint pdf of some continuous bivariate random vector (X, Y)

Review

Definition 4.2.1: Let (X, Y) be a discrete bivariate random vector with joint pmf $f(x, y)$ and marginal pmfs $f_X(x)$ and $f_Y(y)$. For any x such that $P(X = x) = f_X(x) > 0$, the **conditional pmf of Y given that $X = x$** is the function of y denoted by $f(y|x)$ and defined by

$$f(y|x) = P(Y = y|X = x) = \frac{f(x,y)}{f_X(x)}$$

For any y such that $P(Y = y) = f_Y(y) > 0$, the **conditional pmf of X given that $Y = y$** is the function of x denoted by $f(x|y)$ and defined by

$$f(x|y) = P(X = x|Y = y) = \frac{f(x,y)}{f_Y(y)}$$

Review

- Define **conditional expected value** of $g(Y)$ given $X = x$ as

$$\mathrm{E}\{g(Y)|x\} = \sum_y g(y)f(y|x)$$

Review

Definition 4.2.3: Let (X, Y) be a continuous bivariate random vector with joint pdf $f(x, y)$ and marginal pdfs $f_X(x)$ and $f_Y(y)$. For any x such that $f_X(x) > 0$, the **conditional pdf of Y given that X** = x is the function of y denoted by $f(y|x)$ and defined by

$$f(y|x) = \frac{f(x, y)}{f_X(x)}$$

For any y such that $f_Y(y) > 0$, the **conditional pdf of X given that Y** = y is the function of x denoted by $f(x|y)$ and defined by

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Review

- Define **conditional expected value** of $g(Y)$ given $X = x$ as

$$\mathsf{E}\{g(Y)|x\} = \int_{-\infty}^{\infty} g(y)f(y|x)dy$$

Review

- The variance of the probability distribution described by $f(y|x)$, denoted by $\text{Var}(Y|x)$, is called the **conditional variance** of Y given $X = x$

$$\text{Var}(Y|x) = \text{E}(Y^2|x) - \{\text{E}(Y|x)\}^2$$

Review

Definition 4.2.5: Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$ and marginal pdfs or pmfs $f_X(x)$ and $f_Y(y)$. Then X and Y are called **independent random variables** if for every $x \in R$ and $y \in R$,

$$f(x, y) = f_X(x)f_Y(y)$$

Review

Lemma 4.2.7: Let (X, Y) be a bivariate random vector with joint pdf or pmf $f(x, y)$. Then X and Y are **independent random variables** if and only if there exist functions $g(x)$ and $h(y)$ such that, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$,

$$f(x, y) = g(x)h(y)$$

Review

Theorem 4.2.10: Let X and Y be independent random variables.

- For any $A \quad \boldsymbol{R}$ and $B \quad \boldsymbol{R}$,

$$P(X \quad A, Y \quad B) = P(X \quad A)P(Y \quad B)$$

  that is, the events $\{X \quad A\}$ and $\{Y \quad B\}$ are **independent events**

- Let $g(x)$ be a function only of x and $h(y)$ be a function only of y. Then

$$\mathrm{E}\{g(X)h(Y)\} = \mathrm{E}\{g(X)\} \cdot \mathrm{E}\{h(Y)\}$$

Review

Theorem 4.2.12: Let X and Y be independent random variables with moment generating functions $M_X(t)$ and $M_Y(t)$. Then the moment generating function of the random variable $Z = X + Y$ is given by

$$M_Z(t) = M_X(t)M_Y(t)$$

Review

Theorem 4.2.14: Let X $\quad$ n(μ, σ^2) and Y $\quad$ n(γ, τ^2) be independent normal random variables. Then the random variable $Z = X + Y$ has a n$(\mu + \gamma, \sigma^2 + \tau^2)$ distribution

Review

- Bivariate transformations (see previous notes for details)
  - **Discrete case**

$$f_{U,V}(u,v) = P(U = u, V = v) = P((X,Y) \quad A_{uv}) = \sum_{(x,y) \quad A_{uv}} f_{X,Y}(x,y)$$

  - **Continous case (assuming 1 to 1 transformation)**

$$f_{U,V}(u,v) = \begin{cases} f_{X,Y}(h_1(u,v), h_2(u,v))|J| & (u,v) \quad B \\ 0 & otherwise \end{cases}$$

  - **Continous case (not 1 to 1)**

$$f_{U,V}(u,v) = \sum_{i=1}^{k} f_{X,Y}(h_{1i}(u,v), h_{2i}(u,v))|J_i|$$

Review

Theorem 4.3.5: Let X and Y be independent random variables. Let $g(X)$ be a function only of X and $h(Y)$ be a function only of Y. Then the random variables $U = g(X)$ and $V = h(Y)$ are independent

Review

Theorem 4.4.3: If X and Y are any two random variables, then

$$\mathrm{E}(X) = \mathrm{E}\{\mathrm{E}(X|Y)\}$$

provided that the expectations exist

Theorem 4.4.7 (Conditional variance identity): For any two random variables X and Y,

$$\mathrm{Var}(X) = \mathrm{E}\{\mathrm{Var}(X|Y)\} + \mathrm{Var}\{\mathrm{E}(X|Y)\}$$

provided that the expectations exist

Review

Definition 4.5.1: The **covariance** of X and Y is the number defined by

$$\text{Cov}(X,Y) = \text{E}\{(X-\mu_X)(Y-\mu_Y)\}$$

Theorem 4.5.3: $\text{Cov}(X,Y) = \text{E}(XY) - \mu_X\mu_Y$

Definition 4.5.2: The **correlation** of X and Y is the number defined by

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

The value $\rho_{X,Y}$ is also called the **correlation coefficient**

Review

Theorem 4.5.5: If X and Y are independent random variables, then Cov$(X, Y) = 0$ and $\rho_{X,Y} = 0$

- Independence implies 0 covariance, but not the other way around

Review

Theorem 4.5.6: If X and Y are any two random variables and a and b are any two constants, then

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y)$$

If X and Y are independent random variables, then

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

Review

Theorem 4.5.7: For any random variables X and Y,

- $-1 \leq \rho_{XY} \leq 1$
- $|\rho_{XY}| = 1$ if and only if there exist numbers $a = 0$ and b such that $P(Y = aX + b) = 1$. If $\rho_{XY} = 1$, then $a > 0$, and if $\rho_{XY} = -1$, then $a < 0$

Review

- Multivariate distributions (see previous notes for details)
  - joint distribution
  - marginal distribution
  - conditional distribution
  - expected value

Review

Definition 4.6.5: Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random vectors with joint pdf or pmf $f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Let $f_{\boldsymbol{X}_i}(\boldsymbol{x}_i)$ denote the marginal pdf or pmf of $\boldsymbol{X}_i$. Then $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are called **mutually independent random vectors** if, for every $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$,

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = f_{\boldsymbol{X}_1}(\boldsymbol{x}_1) \cdots f_{\boldsymbol{X}_n}(\boldsymbol{x}_n) = \prod_{i=1}^{n} f_{\boldsymbol{X}_i}(\boldsymbol{x}_i)$$

- If X_is are all one-dimensional, then $X_1, \ldots, X_n$ are called **mutually independent random variables**

Review

Theorem 4.6.6 (Generalization of Theorem 4.2.10): Let $X_1, \ldots, X_n$ be **mutually independent random variables**. Let $g_1, \ldots, g_n$ be real-valued functions such that $g_i(x_i)$ is a function only of x_i, $i = 1, \ldots, n$. Then

$$\mathrm{E}\{g_1(X_1) \cdots g_n(X_n)\} = \mathrm{E}\{g_1(X_1)\} \cdots \mathrm{E}\{g_n(X_n)\}$$

Review

Theorem 4.6.7 (Generalization of Theorem 4.2.12): Let $X_1, \dots, X_n$ be **mutually independent random variables** with mgfs $M_{X_1}(t), \dots, M_{X_n}(t)$. Let $Z = X_1 + \cdots + X_n$. Then the mgf of Z is

$$M_Z(t) = M_{X_1}(t) \cdots M_{X_n}(t)$$

In particular, if $X_1, \dots, X_n$ all have the same distribution with mgf $M_X(t)$, then

$$M_Z(t) = \{M_X(t)\}^n$$

Review

Theorem 4.6.11 (Generalization of Lemma 4.2.7): Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random vectors. Then $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are mutually independent random vectors if and only if there exist functions $g_i(x_i)$, $i = 1, \ldots, n$, such that the joint pdf or pmf of $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ can be written as

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = g_1(\boldsymbol{x}_1) \cdots g_n(\boldsymbol{x}_n)$$

Review

Theorem 4.6.12 (Generalization of Theorem 4.3.5): Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be independent random vectors. Let $g_i(_{xi})$ be a function only of $\boldsymbol{x}_i$, $i = 1, \ldots, n$. Then the random variables $U_i = g_i(\boldsymbol{X}_i)$, $i = 1, \ldots, n$, are mutually independent

Review

- Chebyshev's inequality
- Markov inequality
- Normal tail probability
- Hölder's inequality
- Cauchy-Schwarz inequality
- Covariance inequality
- Minkowski's inequality
- Jensen's inequality
- Inequality for means

Review

Definition 5.1.1: The random variables $X_1, \ldots, X_n$ are called a **random sample** of size n from the population $f(x)$ if $X_1, \ldots, X_n$ are mutually independent random variables and the marginal pdf or pmf of each X_i is the same function $f(x)$. Alternatively, $X_1, \ldots, X_n$ are called **independent and identically distributed random variables** with pdf or pmf $f(x)$. This is commonly abbreviated to **iid** random variables

Review

Definition 5.2.1: Let $X_1, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued function whose domain includes the sample space of $(X_1, \ldots, X_n)$. Then the random variable or random vector $Y = T(X_1, \ldots, X_n)$ is called a **statistic**. The probability distribution of a statistic Y is called the **sampling distribution of** Y

Review

Definition 5.2.2: The **sample mean** is the arithmetic average of the values in a random sample. It is usually denoted by

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Definition 5.2.3: The **sample variance** is the statistic defined by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

The **sample standard deviation** is the statistic defined by $S = \sqrt{S^2}$

Review

Theorem 5.2.6: Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then

- $\mathrm{E}(\bar{X}) = \mu$ (unbiasedness)
- $\mathrm{Var}(\bar{X}) = \dfrac{\sigma^2}{n}$
- $\mathrm{E}(S^2) = \sigma^2$ (unbiasedness)

Review

- **Important factoids** Let $X_1, \dots, X_n$ be random variables whose expectations and variances exist
  - $\mathrm{E}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{E}(X_i)$
  - $\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i=j} \mathrm{Cov}(X_i, X_j)$
- Note for a random sample,

$$\mathrm{Var}(X_1 + \cdots + X_n) = n\mathrm{Var}(X_1)$$

- Note for a previous example of sampling without replacement,

$$\mathrm{Var}(X_1 + \cdots + X_n) = n\mathrm{Var}(X_1) + n(n-1)\mathrm{Cov}(X_1, X_2)$$

Review

Theorem 5.2.7: Let $X_1, \ldots, X_n$ be a random sample from a population with mgf $M_X(t)$. Then the mgf of the sample mean is

$$M_{\bar{X}}(t) = \{M_X(t/n)\}^n$$

Review

- When the mgf of $\bar{X}$ is not recognizable, or the population mgf does not exist, the transformation method might be used. In such cases, the following **convolution formula** is useful

Theorem 5.2.9: If X and Y are **independent**, continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, then the pdf of $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(w) f_Y(z - w) dw$$

Theorem 5.3.1: Let $X_1, \ldots, X_n$ be a random sample from a n(μ, σ^2) distribution, and let $\bar{X} = (1/n) \sum_{i=1}^{n} X_i$ and $S^2 = \{1/(n-1)\} \sum_{i=1}^{n} (X_i - \bar{X})^2$. Then

- $\bar{X}$ and S^2 are independent random variables
- $\bar{X} \quad$ n$(\mu, \sigma^2/n)$
- $(n-1)S^2/\sigma^2 \quad \chi^2_{n-1}$

Review

Definition 5.4.1: The **order statistics** of a random sample $X_1, \ldots, X_n$ are the sample values placed in ascending order. They are denoted by $X_{(1)}, \ldots, X_{(n)}$

Review

Theorem 5.4.4: Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the pdf of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)\{F_X(x)\}^{j-1}\{1 - F_X(x)\}^{n-j}$$

- Proof: $F_{X_{(j)}}(x) = \sum_{k=j}^{n} \binom{n}{k} \{F_X(x)\}^k \{1 - F_X(x)\}^{n-k}$ and then differentiate

Review

Theorem 5.4.6: Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of a random sample, $X_1, \ldots, X_n$, from a continuous population with cdf $F_X(x)$ and pdf $f_X(x)$. Then the joint pdf of $X_{(i)}$ and $X_{(j)}$, $1 \le i < j \le n$, is

$$\begin{aligned} & f_{X_{(i)},X_{(j)}}(u,v) \\ = \; & \frac{n!}{(i-1)!(j-1-i)!(n-j)!} f_X(u) f_X(v) \{F_X(u)\}^{i-1} \\ & \times \{F_X(v) - F_X(u)\}^{j-1-i} \{1 - F_X(v)\}^{n-j} \end{aligned}$$

for $-\infty < u < v < \infty$

Review

$X_n \xrightarrow{p} X$: A sequence of random variables, $X_1, X_2, \ldots$, **converges in probability** to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1$$

$X_n \xrightarrow{a.s.} X$: A sequence of random variables, $X_1, X_2, \ldots$, **converges almost surely** to a random variable X if, for every $\epsilon > 0$,

$$P(\lim_{n \to \infty} |X_n - X| < \epsilon) = 1$$

Review

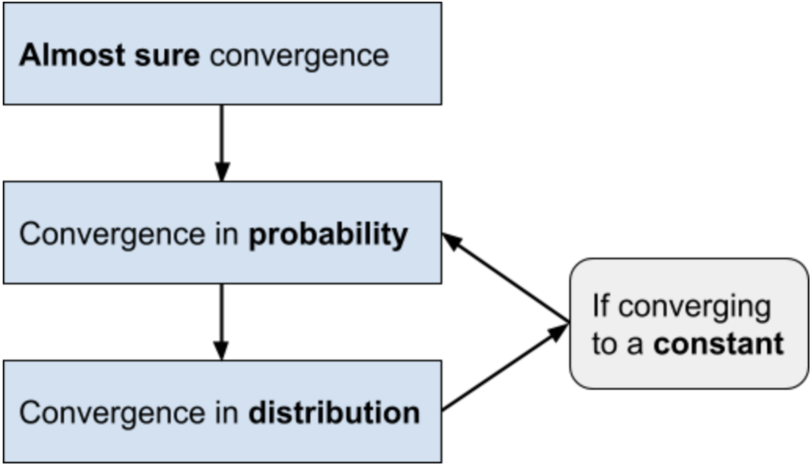
Definition 5.5.10: A sequence of random variables, $X_1, X_2, \ldots$ **converges in distribution** to a random variable X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous. We also write this

$$X_n \xrightarrow{d} X$$

Review


Almost sure convergence
Convergence in **probability**
Convergence in **distribution**
If converging
to a **constant**

Review

Theorem 5.5.2 (Weak Law of Large Numbers): Let $X_1, X_2, \ldots$be iid random variables with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Then, for every $\epsilon > 0$,

$$\lim_{n \to \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

That is, $\bar{X}_n$ **converges in probability** to μ:

$$\bar{X}_n \xrightarrow{p} \mu$$

We say that $\bar{X}_n$ is **consistent** for μ

Review

Theorem 5.5.9 (Strong Law of Large Numbers): Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Then, for every $\epsilon > 0$,

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1$$

That is,

$$\bar{X}_n \xrightarrow{a.s.} \mu$$

Review

Theorem 5.5.14 (Central Limit Theorem): Let $X_1, X_2, \ldots$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (i.e. $M_{Xi}(t)$ exists for $|t| < h$, for some positive h). Let $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 > 0$ (Both μ and σ^2 are finite because the mgf exists). Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any $x \in (-\infty, \infty)$,

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

That is,

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} \mathrm{n}(0,1)$$

Theorem 5.5.15 (Stronger form of the CLT): Let $X_1, X_2, \ldots$ be a sequence of iid random variables with $\mathrm{E}(X_i) = \mu$ and $0 < \mathrm{Var}(X_i) = \sigma^2 < \infty$. Define $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any $x \in (-\infty, \infty)$,

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

That is,

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} \mathrm{n}(0,1)$$

Review

- **Continuous mapping theorem / Mann-Wald mapping theorem**
  Suppose that $X_1, X_2, \ldots$is a sequence of random variables and f : $R \to R$ is a Borel function (includes continuous functions) whose set D of discontinuities is such that $\omega : X(\omega)$ D F and $P(X$ $D) = 0$. If X_n converges to X either
  - almost surely
  - in probability, or
  - in distribution

  Then $g(X_n)$ converges to $g(X)$ in the same sense

Review

Theorem 5.5.17 (Slutsky's Theorem): If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} a$, where a is a constant, then

- $X_n Y_n \xrightarrow{d} aX$
- $X_n + Y_n \xrightarrow{d} X + a$

Review

Theorem 5.5.24 (Delta Method): Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \xrightarrow{d} \mathrm{n}(0, \sigma^2)$. For a given function g and a specific value of θ, suppose that $g\ (\theta)$ exists and is not 0. Then

$$\sqrt{n}\{g(Y_n) - g(\theta)\} \xrightarrow{d} \mathrm{n}(0, \sigma^2\{g\ (\theta)\}^2)$$

Theorem 5.5.26 (Second-order Delta Method): Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \xrightarrow{d} \mathrm{n}(0, \sigma^2)$. For a given function g and a specific value of θ, suppose that $g\ (\theta) = 0$ and $g\ \ (\theta)$ exists and is not 0. Then

$$n\{g(Y_n) - g(\theta)\} \xrightarrow{d} \sigma^2 \left\{\frac{g\ \ (\theta)}{2}\right\} \chi_1^2$$

Review

- Let $\boldsymbol{T} = (T_1, \ldots, T_k)^T$ with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\Sigma_{\boldsymbol{T}}$. Let $\boldsymbol{Q} = \boldsymbol{G}(\boldsymbol{T}) = (g_1(\boldsymbol{T}), \ldots, g_m(\boldsymbol{T}))^T$. Then

$$\mathsf{E}(\boldsymbol{Q}) \approx \boldsymbol{G}(\boldsymbol{\mu}) \quad \text{and} \quad \mathsf{Var}(\boldsymbol{Q}) \approx H(\boldsymbol{\mu})\Sigma_{\boldsymbol{T}} H(\boldsymbol{\mu})^T$$

where $H(\boldsymbol{\mu}) = H(\boldsymbol{t})|_{t=\boldsymbol{\mu}}$ and

$$H(\boldsymbol{t}) = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{t})}{\partial t_1} & \cdots & \frac{\partial g_1(\boldsymbol{t})}{\partial t_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(\boldsymbol{t})}{\partial t_1} & \cdots & \frac{\partial g_m(\boldsymbol{t})}{\partial t_k} \end{bmatrix}$$

Review

- Final exam focuses on materials after the midterm (after Section 3.3)
- A couple of notes $(1/3)$
  - $\mathrm{E}(\cdot)$
    - is sum or integration
      includes mean, variance, moments and mgfs
    - handy calculation of $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ using hierarchy
  - Derive marginal and conditional pdfs/pmfs from a joint pdf/pmf
  - Univariate, bivariate, or multivariate transformations
    - discrete case or continuous case?
    - do we have 1 to 1 transformation?

Review

- A couple of notes (2/3)
  - mgfs can be used to
    - derive moments
    - derive variance through 1st and 2nd moments
    - name/recognize a distribution (note: simplified derivation for sum of independent r.v.s)
    - examine convergence in distribution
  - Independence of r.v.s
    - is verified when both joint density and support region factor
    - implies the independence of events and functions of r.v.s.
    - implies that, expectation of product = product of expectation
    - implies that covariance and correlation = 0; the opposite direction is true for normal r.v.s but does not generally hold

Review

- A couple of notes (3/3)
  - Relations between r.v.s (e.g. square of standard normal is chi-square)
  - Random sample means iid
  - Sample mean and sample variance are unbiased. More properties when assuming normal distribution
  - Order statistics (min, max and more)
  - Probabilistic inequalities and applications
  - Three modes of convergence and their relations
  - How to apply LLN, CLT and Delta method and when?

Review

- Always double check assumptions when applying any theorems or properties (e.g. $M_{X+Y}(t) = M_X(t)M_Y(t)$ require X and Y be independent)
- Familiarity with common distributions and their relations is useful and sometimes can save time
- Useful tools: proof by contradiction, induction, recursive relation, recognize a kernel, integration (dxdy or dydx, polar coordinates), geometric argument (e.g. find the area), inequalities, Taylor expansion, etc.
- Pay attention to support region
- Go back to definitions if no clue; they are fundamental and often the first step of solutions